Sequential three-way decision and granulation for cost-sensitive face recognition

Huaxiong Li\textsuperscript{a,}\textsuperscript{*}, Libo Zhang\textsuperscript{a}, Bing Huang\textsuperscript{b}, Xianzhong Zhou\textsuperscript{a}

\textsuperscript{a} Department of Control and Systems Engineering, School of Management and Engineering, Nanjing University, Nanjing 210093, China
\textsuperscript{b} School of Technology, Nanjing Audit University, Nanjing, Jiangsu 211815, China

Article info

Article history:
Received 25 January 2015
Revised 29 July 2015
Accepted 30 July 2015
Available online 6 August 2015

Keywords:
Three-way decisions
Decision-theoretic rough sets
Granular computing
Cost-sensitive learning
Face recognition

Abstract

Many previous studies on face recognition attempted to seek a precise classifier to achieve a low misclassification error, which is based on an assumption that all misclassification costs are the same. In many real-world scenarios, however, this assumption is not reasonable due to the imbalanced misclassification cost and insufficient high-quality facial image information. To address this issue, we propose a sequential three-way decision method for cost-sensitive face recognition. The proposed method is based on a formal description of granular computing. It develops a sequential strategy in a decision process. In each decision step, it seeks a decision which minimizes the misclassification cost rather than misclassification error, and it incorporates the boundary decision into the decision set such that a delayed decision can be made if available high-quality facial image information is insufficient for a precise decision. To describe the granular information of the facial image in three-way decision steps, we develop a series of image granulation methods based on two-dimensional subspace projection methods including 2DPCA, 2DLDA and 2DLPP. The sequential three-way decisions and granulation methods present an applicable simulation on human decisions in face recognition, which simulate a sequential decision strategy from rough granule to precise granule. The experiments were conducted on two popular facial image database, which validated the effectiveness of the proposed methods.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Computer-aided face recognition systems have received much attention over the past few decades. Numerous face recognition techniques were developed [1–7]. Such techniques are capable of cooperating with human users in many applications, e.g., entrance-guard system, citizenship identification system, computer access control, and electronic surveillance [8]. Most previous studies on face recognition attempted to achieve a low misclassification error, i.e., the higher the recognition accuracy, the better the method [8]. In many real-world scenarios, however, this evaluation is not reasonable. The reasons lie in two aspects. First, in real world face recognition, different kinds of misclassifications will lead to different costs. For example, consider an entrance guard system via face recognition techniques, the cost of misrecognizing an impostor as an office member would be much larger than that of misrecognizing an office member as an impostor, since the former may cause much worse consequence than the latter [9]. The examples indicate that the misclassification costs are usually quite different in reality, and simply minimizing the misclassification error rate may not be a good objective for face recognition systems. It is necessary to develop face recognition techniques that can deal with cost-sensitive classification problem [10]. Second, a high-quality image data set is necessary for high-precise recognition. However, in many real-world applications, it is difficult to obtain high-quality images, or it may associate with extra costs to get these high-quality images. In the case that we only have some low-quality images, a delayed decision may be a better choice than an immediate decision, since the cost of making a delayed decision is lower than that of making a certain decision. However, with the increasing of available high-quality images, the cost of deciding a delayed decision may be higher than that of deciding positive or negative decision, thus many delayed decision will be definitely decided to positive or negative region. These decisions lead to a three-way decision strategy [11–15].

The methodology of three-way decisions are commonly used in human daily decision making [14] and widely applied in many theoretic fields, such as management sciences [16], social judgement theory [17], and hypothesis testing in statistics [18]. More practical applications of three-way decisions were reported in numerous...
fields, including, medical decision making [19–22], peer review process [23], government and investment decision [24,25], Email spam filtering [26,27], text classification [28], clustering analysis [29–31], semi-supervised learning [32], knowledge granulation [33], and multi-agent decision [34]. Although the essential ideas of three-way decisions are frequently used in both theoretic and practical research fields, there are few literatures concerning a unified formal description on three-way decisions [14]. To address this issue, Yao presented a general overview on existing three-way decisions researches, and extended the rough sets-based three-way decisions to a much wider frontier, which outlines a unified theory of three-way decisions [14]. It is emphasized that the evaluation of acceptance and rejection is a primitive notion for characterizing the satisfiability of objects, which can be regarded as a key concept of three-way decisions.

Yao [14] and Hu [35] presented a comprehensive survey on three-way decisions. According to the methods used for evaluating acceptance and rejection, they classified the three-way decisions into three categories: three-way decisions with a pair of postet-based evaluations, three-way decisions with one postet-based evaluation, and three-way decisions with one postet-based evaluation [14,35]. From a theoretical point of view, such evaluations can be determined by a set of criteria, such as costs, benefits, and degrees of desirability [14]. Further studies on the criteria of evaluation may be a meaningful research direction in three-way decision field. Given a certain criterion of evaluation, the objective of three-way decisions can be formulated as an optimization problem to minimize or maximizing the evaluation values. Therefore, different kinds criteria lead to different three-way decision models. For example, if one considers the cost or risk as a criterion of evaluation, and discusses the uncertainty of decision from a rough set point of view, then the purpose of three-way decisions is to minimize the overall costs among three decision regions, which leads to a rough set-based three-way decisions [12,15,36–38]. Decision-theoretic rough sets model is a typical rough set-based three-way decisions model, in which a quantitative inclusion of an equivalence class is used for evaluating acceptance and rejection [36]. It can be regarded as a generalized rough set-based three-way decisions model [12]. If one considers a qualitative inclusion of an equivalence class for evaluating acceptance and rejection, then a Pawlak rough set-based three-way decisions model is derived, which can be viewed as an extreme case of decision-theoretic rough set-based three-way decision model [14]. Moreover, if one considers methods to evaluate the acceptance and rejection, or uses further approaches to represent the uncertain and insufficient information of a concept, then some new models of three-way decisions can be derived, such as interval set-based models [14,39,40], fuzzy set-based models [41], and shadowed sets-based models [42,43]. In this paper, we focus on the rough sets based three-way decisions models, especially on the decision-theoretic rough sets-based models, since it takes the cost as an evaluation for acceptance and rejection, which is suitable for the cost-sensitive classification and decision problems.

Recently, rough sets based three-way decisions models received much attention in decision-making and artificial intelligence research fields [13,44,45,24,27,29,46]. It presents a well semantics explanation on how to classify an instance into positive, negative or boundary regions based on cost-sensitive classification strategy. A core content of rough sets based three-way decisions is to introduce boundary decision to traditional two-way decisions (positive and negative decisions) [13], and seek an optimal decision from the three regions by minimizing the decision cost. It is evident that, in some cases, the boundary decision may achieve a lower cost than positive and negative decisions do if available information for immediate decision is insufficient, which is consistent with human decision process [47]. For example in criminal investigation, it is difficult for a police officer to immediately identify a criminal from her/his first impression on facial features, although in memory she/he has recognized the person as a criminal. In this case, neither positive (criminal) nor negative (innocent) decision is immediately decided. The officer will delay the decision and collect more information for further decision, since this boundary decision will cause the lowest mistake cost. This decision method takes a typical three-way decision strategy.

If one considers the applications of rough set-based three-way decision in cost-sensitive face recognition, there are still two important problems need to be further studied. First, the boundary decision presents a delayed solution when available information is insufficient. However, if available information increases, the previous boundary decisions may be converted to positive or negative decisions. How to describe such dynamic sequential decision process is meaningful for applications. Second, how to evaluate the available information or granular feature of a facial image should be considered. Human vision can easily deal with face recognition from rough granule to precise granule. While for computer vision, it is difficult to describe the granular structure of an image and evaluate the available granular information for decision. To address these two issues, we propose a dynamic sequential three-way decision method for cost-sensitive face recognition. A series two-dimension subspace feature extraction methods including 2DPCA [4,48], 2DLDA [49] and 2DLP [50] are introduced to describe the available information in facial image. The principal components induced from the subspace projections are adopted as the representations of the image granular structures. The proposed sequential three-way decision takes a dynamic strategy for face recognition. In the case that available information is insufficient, i.e., rough granule case, some images are wrong recognized or decided as boundary decisions, neither positive nor negative decision, thus the decision costs stay in a high level. With the increasing of available information, the decision costs decrease since the decisions are more precise and some boundary images are converted to certain positive or negative region. Such sequential three-way decision presents a simulation on human dynamic decision process in cost-sensitive face recognition.

This paper is an extended version of the paper [45]. The remainder of this paper is organized as follows. In Section 2, we briefly introduce some related work on three-way decision. In Section 3, we formulate the cost-sensitive face recognition problems. In Section 4, we propose a dynamic cost-sensitive sequential decision method. Section 5 presents two-dimensional subspace methods for facial image granulation, which are used to extract granular features from the images, such that cost-sensitive sequential decision can be conducted on the granular features. Section 6 reports the experimental results and analysis. Finally, in Section 7, we conclude the paper.

2. Related work

The theories and methods on three-way decisions have been developed for a long history, which can be traced back to some early studies on the social judgement theory [14,17]. Many recent works further developed the theories and applications of three-way decisions [27,29,35,46,51,52]. Hu presented a comprehensive survey on three-way decisions in [35]. In general, there are two main aspects in the three-way decisions researches. The first aspect is the researches on the background of three-way decisions [27,29,35,46,51,52], and the second aspect is the theoretical framework researches on three-way decisions [14,35].

The first aspect mainly focuses on the extension researches of rough sets, including two categories works: the extensions of Pawlak rough sets to new rough set models, and the extensions
of one granular to multi-granulation [35]. The first category mainly concerns the extension of the Pawlak rough sets to generalized rough sets, such as decision-theoretic rough sets (DTRS) [27,29,35,46,51–53], stochastic DTRS [54], variable precision rough sets (VPRS) [55], Bayesian rough sets (BRS) [56], game-theoretic rough sets(GTRS), fuzzy rough sets [41], interval-valued fuzzy rough sets (IVFRS) [57–59], interval sets [14,39,40], and dominance-based fuzzy rough sets [60]. The second category focuses on the extensions from one granular to multi-granulation, such as multi-granulation rough sets (MGRS) [61], updated multi-granulation rough approximations [62], multi-granulation decision-theoretic rough sets [63], multi-granulation rough sets based covering [64], and neighborhood-based multi-granulation rough sets (NMGRS) [65,66]. Another extension work on three-way decisions is to extend static three-way decision to the dynamic version. For example, Yao first proposed a sequential three-way decision model from granular computing view point [67,15]. Li proposed a dynamic sequential three-way decision model based on DTRS [44]. Liu et al. developed a dynamic three-way decision model based on DTRS [68]. Luo et al. discussed the dynamic maintenance principles of three-way decision rules [69].

The second aspect is theoretical framework researches on three-way decisions. Yao presented a general outline of the three-way decision theory. It mainly contains value domain of evaluation functions [14], construction and interpretation of evaluation functions [12,13] and the models of three-way decisions [14]. From this point of view, the evaluations in three-way decisions can be determined by a set of criteria including costs, benefits, and degrees of desirability, in which the costs is a widely-used evaluation function. Cost function analysis is important theoretic topic in rough set-based decision. In the context of cost-sensitive face recognition and three-way decision, adecision result can be represented as

\[ A = \begin{bmatrix} A_1 & A_2 & \cdots & A_C \end{bmatrix}, \quad N_t = \begin{bmatrix} 1 & 2 & \cdots & c \end{bmatrix}, \quad m \times n \quad \text{image matrix}, \quad N_i \quad \text{total number of images}. \]

The class label of \( A_i \) is denoted as \( i \in \{1, 2, \ldots, c + 1\} \), where the former \( c \) labels are gallery subjects, and the last one denotes the impostor class. For simplicity, we use \( P \) represents all gallery subjects label (e.g. family members or employees of a company [9]), and \( N \) denotes the label of impostor. In the context of cost-sensitive face recognition and three-way decision, a decision result can be represented as \( D = \{q \_a, q \_b, q \_g\} \), which denotes classifying a facial image as gallery subject (positive decision), impostor (negative decision), and boundary (delayed decision), respectively. The costs of different decisions are discriminatingly treated. Therefore, all recognition decisions fall into six categories:

1. **True acceptance**: correctly classifying a gallery subject.
2. **True rejection**: correctly classifying an impostor.
3. **False acceptance**: misclassifying an impostor as a gallery subject.
4. **False rejection**: misclassifying a gallery subject as an impostor.

In addition, the applications of rough set-based three-way decisions were widely concerned in many fields and disciplines. Besides the applications work mentioned in last section, more practical applications can be found in recent literatures. For example, Yu et al. proposed a series three-way clustering methods for density-based overlapping clustering [29], incomplete data clustering [30], incremental clustering [31], and tree-based incremental overlapping clustering [80]. Zhou and Jia discussed the applications of three-way decision in E-mail spam filtering [26,27]. Liu et al. applied three-way decision into investment decision [24] and government decision [25]. Miao et al. developed three-way decision methods for the applications in text classification [28], semi-supervised learning [32], and knowledge granulation [33]. Yang and Yao investigated the applications of three-way decision in multi-agent decision [34]. Zhang et al. introduced three-way decision to face recognition and proposed a cost-sensitive sequential three-way decision method for face recognition [45]. Zhang and Min combined three-way decision with random forest, and applied three-way decision into recommender systems [81]. The above mentioned works enrich the theoretic foundation of rough set-based three-way decisions, and indicate that it is suitable for many practical decision problems.

In face recognition research field, the importance of feature extraction cannot be overestimated. Subspace projections are widely used in feature extraction fields. Turk and Pentland first introduced PCA to feature extraction and developed the eigenfaces for facial face recognition [3]. It is an unsupervised method which may lose the label information. Belhumeur et al. introduced LDA to subspace projection and proposed the fisherfaces for facial feature extraction, which takes the label information into consideration [1]. He et al. proposed a LPP subspace projection method, and developed the laplacianfaces for face recognition [2]. The aforementioned three subspace projections are all 1D vector-based methods, which requires to convert 2D images to 1D vectors. It may cause expensive time consumption and high space complexity. Besides, the underlying data structures of an image may be lost in the process of converting 2D images to 1D vectors. To address these issues, Yang first proposed an 2D image-based PCA (2DPCA) method [4]. Using the similar projection method in 2DPCA, many 2D subspace projections were proposed, including 2DLDA [5], 2DLPP [50,6]. These 2D subspace projection methods can be used to extract facial features that preserve the underlying data structure of the images, and have much low time and space complexities, since they extract the feature directly from 2D images rather than converting the images to 1D vectors. In this paper, we introduce 2DPCA, 2DLDA and 2DLPP to represent the granular feature of an image, and we will prove that, the features extracted by 2D subspace projection methods are in accord with the definition of the granular feature. Thus they provide a feasible method for image granulation, which is a key step for sequential three-way decisions. We will present a formal description on these issues in our following discussions.
Boundary acceptance: classifying a gallery subject after delay.

Boundary rejection: classifying an impostor after delay.

Denote the costs of (1)–(6) as $\lambda_{BP}$, $\lambda_{BN}$, $\lambda_{PN}$, $\lambda_{BP}$, and $\lambda_{BN}$, respectively, and we obtain a cost matrix $(\lambda_{ij})_{2 \times 2}$, where $i \in \{P, B, N\}, j \in \{P, N\}$. The costs of these decisions are not assumed to be equal. Normally the cost for a right decision is less than that of boundary decision (delayed decision), and less than that of wrong decision, thus we have $\lambda_{BP} \leq \lambda_{BN} \leq \lambda_{PN}$ and $\lambda_{BN} \leq \lambda_{BN} \leq \lambda_{BN}$.

As discussed in Section 1, human decisions for face recognition usually takes a sequential and hierarchical strategy, which can be described using granular structure [82,83]. From a granular computing point of view, the features of the facial images in a sequential decision process can be represented using total order relation [15]. For clarity, we present a formal definition on the granular feature and function for an image data set as follows:

**Definition 1.** Let $\mathcal{M} = \{A_1, A_2, \ldots, A_N\}$ be an image data set, where $A_i \in \mathbb{R}^{m \times n}$. Let $F = \{f^1, f^2, \ldots, f^m\}$ be a mapping set, where $f^i : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times l}$ (l = 1, 2, ..., n) is a mapping from the image data space to feature space, which is used as a granular descriptor for facial image. For $\forall A \in \mathcal{M}$, a feature set $B = \{B^1, B^2, \ldots, B^n\} = \{f^1(A), f^2(A), \ldots, f^m(A)\}$ is called as a granular feature set if ($B, \preceq$) is a totally ordered feature set under a totally ordered relation $\preceq$, i.e., $B^1 \preceq B^2 \cdots \preceq B^n$, and we call $F$ and $l$ as granular function and granular index, respectively.

Different from traditional cost-blind classification and decision methods, the purpose of the sequential cost-sensitive face recognition is to find an appropriate decision from the available information. It may cause some decisions, the available information is rarely sufficient. It may cause some costs to acquire new information resource [47]. While new information is available, a more precise decision may be decided. The process will continue until a satisfying decision precision is achieved. Such process forms a sequential three-way decision. It presents a simulation on the human decision processes: from a rough granularity to a precise granularity, which are frequently used in real-world applications. We will present a mathematic description on the sequential decision in this section. First, we present a formal definition on sequential three-way decision from the totally ordered relation point of view as follows.

**Definition 2.** Let $\mathcal{M} = \{A_1, A_2, \ldots, A_N\}$ be an image data set and $D = \{d_p, d_B, d_N\}$ be a decision set, where $A_i \in \mathbb{R}^{m \times n}$. Let $C = \{B^1, B^2, \ldots, B^n\}$ be a granular feature set of an image $A \in \mathcal{M}$. Suppose $C(\preceq) \preceq$ is a totally ordered feature set under a totally ordered relation $\preceq$, i.e., $B^1 \preceq B^2 \cdots \preceq B^n$, and cost $(d(B))$ denotes the cost of deciding $B^j$ as $d$, then the following series is called sequential three-way decision:

$$SD = (SD_1, SD_2, \ldots, SD_5, SD_6) = (\phi^*(B^1), \phi^*(B^2), \ldots, \phi^*(B^j), \ldots, \phi^*(B^n)).$$

where $\phi^*(B^j)$ is the minimum cost decision of the $j$-th step, i.e., $\phi^*(B^j) = \arg \min_{d(B^j)} \text{cost}(d(B^j)).$

For a cost-sensitive sequential three-way decision $SD$, the l-th step decision $\phi^*(B^j)$ is acquired by minimizing the decision cost according to the l-order granular feature information, thus we can solve the minimum decision using an optimization method. Considering an l-th decision on an l-order granular feature $B^l$, the decision costs $(\alpha, \beta)$, $i \in \{P, B, N\}$ can be computed based on Bayesian decision procedure [84,12], then we have:

$$\text{cost}(\alpha(B^l), \beta(B^l)) = \lambda_{BP} \text{Pr}(P(B^l)) + \lambda_{BN} \text{Pr}(N(B^l)),$$

$$\text{cost}(\alpha(B^l), \beta(B^l)) = \lambda_{BN} \text{Pr}(N(B^l)) + \lambda_{BP} \text{Pr}(P(B^l)), (2)$$

$$\text{cost}(\alpha(B^l), \beta(B^l)) = \lambda_{BN} \text{Pr}(N(B^l)) + \lambda_{BN} \text{Pr}(N(B^l)).$$

For any $d \in D$, we compute the decision costs according to (2) and select the decision with the minimum cost as the optimal decision for the l-th step, which is presented in Definition 2 as follows:

$$SD_l = \phi^*(B^l) = \arg \min_{d(B^l)} \text{cost}(d(B^l)).$$

Considering a special case that $(\lambda_{BN} - \lambda_{BN})(\lambda_{BN} - \lambda_{BP}) > (\lambda_{PN} - \lambda_{BN})(\lambda_{BN} - \lambda_{BN})$, which is hold for most cases [36], then the optimization solution of (3) can be obtained according to DTRS model [85,36,37], which is presented as follows:

$$SD_l = \begin{cases} d_p & \text{if } \text{Pr}(P(B^l)) \geq \alpha, \\ d_B & \text{if } \text{Pr}(P(B^l)) \leq \beta, \\ d_N & \text{if } \beta < \text{Pr}(P(B^l)) < \alpha, \end{cases}$$

where the thresholds $\alpha$ and $\beta$ can be computed as follows [36]:

$$\alpha = \frac{\lambda_{BN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \quad \beta = \frac{\lambda_{BN} - \lambda_{BN}}{(\lambda_{BN} - \lambda_{BN}) + (\lambda_{BN} - \lambda_{BP})}. (5)$$

From (3) and (5), we can observe that the optimal decision $SD_l$ in a sequential three-way decisions depend on the loss functions $\lambda_{ij}$ and the conditional probability $\text{Pr}(P(B^l))$. The loss functions can be obtained from experts or learning methods. Once the loss functions are given, the optimal decision is determined by the conditional probability. To estimate the conditional probability $\text{Pr}(P(B^l))$, there is a key point which needs to be further investigated. The granular feature $B^l$ is a granular descriptor of a facial image which satisfies Definition 1. How to get this granular descriptor should be investigated. To address this problem, we introduce a series of two-dimensional subspace granulation to construct the granular features of facial images.

5. Sequential subspace granulation

As mentioned in last section, we should find appropriate methods to represent the granular feature of a facial image, so that the
optimal decision of three-way decision can be determined. In the pattern recognition and machine learning communities, there are many studies concerning on how to represent the intrinsic features of a data set [1–3,86,87], which can be used as the granular descriptors of the data. Many studies demonstrated that the feature of image data set are usually distribute on a low-dimensional subspace or manifold, in which the intrinsic granular structure of the images can be represented [1–3]. In the past decade, two-dimensional subspace projection methods were demonstrated high performance and efficiency for facial image representation [4,6,48–50]. Motivated by the success of two-dimensional subspace projection methods in feature representation, we introduce three two-dimensional subspace methods to obtain the granular feature of facial images, and we will also demonstrate shortly that the two-dimensional subspace projections satisfies the Definition 1 such that they can be viewed as image granulation methods.

5.1. Sequential two-dimensional subspace granulation

Let \( W = [w_1, \ldots, w_n] \) be a square transform matrix, and let \( W_l = [w_1, \ldots, w_l] \) represent a subset transform matrix which consists of the first \( l \) column vectors of \( W \), where \( l \) varies from 1 to \( n \). A feature matrix \( B_1 \) be a granular feature set, and \( F \) is the mapping set such that they can be viewed as image granulation methods.

We can further acquire a feature matrix set of \( A \) if we consider a sequential transform. Let \( W = [w_1, \ldots, w_n] \) be a set of transform matrices, and we get a sequential feature set by projecting \( A \) onto \( [w_1, \ldots, w_n] \), respectively, which is presented as follows:

\[
B_1 = f^P(A) - AW_l, \quad (6)
\]

We argue that the mapping set \( F = \{f^1, f^2, \ldots, f^n\} \) satisfies Definition 1 and thus can be viewed as a granular function, which is formally presented as the following proposition.

**Proposition 1.** Let \( A \) be a data set, and \( W = [w_1, \ldots, w_n] \) be a transform matrix, and let \( W_l = [w_1, \ldots, w_l] \) represent a subset transform matrix which consists of the first \( l \) column vectors of \( W \), where \( l = 1, 2, \ldots, n \). A feature matrix set \( B_1 \) is obtained by projecting \( A \) onto \( W_1 \), respectively, and then the mapping set \( F = \{f^1, f^2, \ldots, f^n\} \) is a granular function, and \( B_1 \) is a granular feature set.

**Proof.** Denote \( B = \{B_1, B_2, \ldots, B_n\} \), i.e., \( B_l = f^P(A) - AW_l, l = 1, \ldots, n \). We define an relation “\( \preceq \)" on feature matrix set \( B \) as follows: for \( \forall B_1, B_2 \in B \), \( B_1 \preceq B_2 \) if and only if all column vectors of \( B_2 \) is the columns of \( B_1 \), i.e., \( B_1 \subseteq B_2 \) (b is a column of the matrix \( B_2 \)), holds that \( B_1 \subseteq B_2 \) (b is a column of the matrix \( B_2 \)). We can prove that \( \preceq \) is a totally ordered relation: (1) Reflexivity: it is straight forward that \( \forall B_1, B_2 \in B_1 \preceq B_2 \) holds. (2) Antisymmetry: \( \forall B_1 \subseteq B_2, B_2 \subseteq B_1 \) and \( \exists p \leq l \) such that \( B_l = AW_p \). Therefore, \( B_1 \subseteq B_2 \) holds. (3) Transitivity: suppose \( B_1 \preceq B_2 \) and \( B_2 \preceq B_3 \), then we have \( B_1 \preceq B_3 \) holds. (4) Totality: \( \forall B_1, B_2 \in W, 3p, q \in \{1, \ldots, n\}, B_1 = [A_{w_1}, \ldots, A_{w_p}], B_2 = [A_{w_1}, \ldots, A_{w_q}] \) if \( p < q \). Then we have \( B_1 \preceq B_2 \), otherwise \( B_1 \preceq B_2 \), thus the totality of \( \preceq \) holds. Therefore, “\( \preceq \)" is a totally ordered relation and \( (B, \preceq) \) is a totally ordered set. According to Definition 1, \( B = \{f^1(A), f^2(A), \ldots, f^n(A)\} \) is a granular feature set, and \( F = \{f^1, f^2, \ldots, f^n\} \) is a granular feature function.

For Proposition 1, we can see that (7) is a granular feature set generated from two-dimensional subspace projection, which represents the granular structure of an image matrix from rough granule to precise granule. According to Definition 2, the three-way decisions on granular feature set \( B \) is a sequential three-way decision. It presents a simulation of human vision decision process from incomplete information to complete information. It is worth noting that (7) presents a general two-dimensional subspace projection for image granulation. However, how to acquire the transform matrix \( W \) should be considered. In the context of cost-sensitive learning, it is important for a granular feature to be well separated and can be precisely classified, such that the misclassification cost may reach a low level. To this end, we introduce three subspace projection methods to acquire the transform matrix \( W \).

5.2. Sequential granulation using 2DPCA

To acquire a sequential granular features of a facial image, we introduce 2DPCA [4] for sequential granular feature extraction, which aims to enlarge the variance of the images in feature space, such that the features are clearly separated and the misclassification cost reach a low level, as mentioned in last section. First, consider a special case of (6), we can project \( A \) onto a single vector \( w \in \mathbb{R}^n \), and we obtain an \( n \times 1 \) feature vector, \( b = Aw \). To maximize the variance of features, we define an objective function \( J_1(w) \) as the feature variance:

\[
J_1(w) = \frac{1}{N_l} \sum_{i=1}^{N_l} \|b_i - \bar{b}\|^2.
\]

where \( b_i \) is the feature vector of a single image \( A_i \), and \( \bar{b} \) is the mean feature vector of all images, and \( \| \cdot \| \) is Euclidean norm. Substituting \( b_i = Aw \) into (8), we can rewrite (8) as follows:

\[
J_1(w) = \frac{1}{N_l} \sum_{i=1}^{N_l} (A_iw - Aw)\top(A_iw - Aw) = \bar{w}^\top S \bar{w}.
\]

where \( S = \frac{1}{N_l} \sum_{i=1}^{N_l} (A_i - \bar{A})\top(A_i - \bar{A}) \) is the covariance matrix of the image set. The optimal projection axis \( \bar{w} \) is obtained by solving the following optimization problem [84]:

\[
\bar{w} = \arg \max_{w} J_1(w) = \arg \max_{w} \bar{w}^\top S \bar{w}.
\]

We enforce a normalization constraint on \( w \), i.e., \( \bar{w}^\top \bar{w} = 1 \) since the length of \( w \) is irrelevant to the projection result [84]. Then the optimal solution of (10) can be obtained by solving an eigenvector equation as follows:

\[
\bar{w} = \lambda \bar{w}.
\]

It can be solved by setting \( w \) equal to the eigenvector that has the largest eigenvalue that is known as the first principal component [84]. In order to acquire a sequential granular feature set, we select \( l \leq l \leq n \) projection axes \( w_1, \ldots, w_l \) that are the orthogonal eigenvectors of \( S \) corresponding to the first eigenvalues of \( S \). Let \( W_{pca} = [w_1, \ldots, w_l] \) and \( f_{pca}(A) = AW_{pca} \), \( l = 1, \ldots, n \), then we obtain a granular feature set \( B \) for an image \( A \) according to (7) as follows:

\[
B_{pca} = \{ f_{pca}(A), \ldots, f_{pca}(A) \} = \{ AW_{pca}^1, AW_{pca}^2, \ldots, AW_{pca}^n \}.
\]

5.3. Sequential granulation using 2DLDA

The granular features extracted by 2DPCA does not involve the label information since it is an unsupervised feature granular
To incorporate the label information, we introduce a supervised two-dimensional subspace projection 2DLD for image granulation. The purpose of 2DLD [49] granulation is to search for a projection axis that can maximize the between-class scatter while simultaneously minimizing the within-class scatter [49]. To this end, we introduce the objective function $J_f(w)$ of 2DLD for image granulation, which is defined as follows:

$$J_f(w) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} ||b_i - b_j||^2,$$

(18)

where $b_i$ is the mean vector of the $i$th class, and $\overline{b}$ is the mean vector of all features, and $c + 1$ is the number of classes (the former $c$ classes are gallery subjects while the last one is impostor). Substituting $b_i = A_iw$ into (18), we can rewrite (18) as follows:

$$J_f(w) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)^T (A_iw - A_jw) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)(A_iw - A_jw)^T.$$

where $W_{ij}$ is a locality measure function [2] to measure the neighbor characteristic of two image features $A_i$ and $A_j$, defined as follows:

$$W_{ij} = \begin{cases} \exp\left(-\frac{||A_i - A_j||^2}{\sigma^2}\right), & \text{if } A_i \in N_k(A_j) \text{ or } A_j \in N_k(A_i), \\ 0, & \text{otherwise.} \end{cases}$$

(19)

Substituting $b_i = A_iw$ into (18), we can rewrite (18) as follows [6]:

$$J_f(w) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} ||A_iw - A_jw||^2 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)(A_iw - A_jw)^T = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)(A_iw - A_jw)^T.$$

(20)

where $A^T = [A_1, \ldots, A_c]$, and $A = [A_1, \ldots, A_N]$. $D_i = \sum_{j=1}^{N_i} W_{ij}$, $D = \text{diag}(D_{11}, \ldots, D_{N1})$, is an $m \times m$ identity matrix, and $W$ is a block matrix, in which the $ith$ row and $jth$ column block is $W_{ij}$, $L = D - W$ is the Laplacian matrix. It is clear that a trivial solution to minimize $J_f(w)$ is $w = 0$. To avoid it, we can impose a constraint as $w^T A D A w = 1$. Therefore, 2DLD-based granular feature extraction aims to find an optimal projection axis $w$ solving the following optimization problem [50]:

$$w = \arg \min_{w^T D A = 1} J_f(w) = \arg \min_{w^T D A = 1} \sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)(A_iw - A_jw)^T.$$

(21)

which can be solved by maximizing the Lagrangian function [84]:

$$\sum_{i=1}^{N} \sum_{j=1}^{N_i} W_{ij} (A_iw - A_jw)(A_iw - A_jw)^T.$$

(22)

The optimal projection axis $X$ can be solved from the following generalized eigenvector equation:

$$A^T L A w = \lambda A D A w.$$

(23)

Since our objective is to minimize the $J_f(w)$, we select $l(1 \leq l \leq n)$ projection axes $w_1, \ldots, w_l$ that are the eigenvectors of $A^T L A$ corresponding to the $l$ smallest eigenvalues [6]. Let $W_{pp} = [w_1, \ldots, w_l]$ and $f_{pp}(A) = AW_{pp}^T$, $l = 1, \ldots, n$. Then, we obtain a granular feature set $B$ for an image $A$ according to (7):

$$B_{pp} = \{f_{pp}(A), f_{pp}(A), \ldots, f_{pp}(A)\} = [AW_{1}^T \ldots AW_{n}^T].$$

(25)

6. Experiments

In this section, experiments were designed to evaluate the performance of the proposed sequential three-way decision and granulation methods on two popular databases AR [88] and PIE [89]. We compared the performance of the proposed methods with those sequential two-way decision, and examine the effectiveness in sequential decision process.

6.1. Face databases

The AR face database contains over 4000 face images of 126 subjects, in which each subject has 26 images with different facial expressions, lighting conditions and occlusions [88]. In our experiments, the main purpose is to test the performances of sequential three-way decision and granulation on cost-sensitive face recognition. Thus we used 14 non-occluded face images of each subject [10], and each image was cropped and resized to 60 × 43 pixels. The PIE face database contains 41,368 face images of 68 subjects as a whole. The face images were taken by 13 synchronized cameras and 21 flashes, with varying pose, illumination, and expression [89]. To test the performance of the algorithm on cost-sensitive face recognition, we selected the front view images [9]. For each subject in PIE database, we used a subset including 49 images. All images were cropped and resized to 60 × 60 pixels. Some sample images of AR and PIE databases used in our experiments are shown in Fig. 1. In addition, all experiments on these databases were implemented on a computer with 2.8G processor and 6G RAM, and programmed in MATLAB (version R2013a).
6.2. Sequential granulation on facial images

An experiment was performed to test the image sequential granulation effectiveness using three two-dimensional subspace methods. Since we consider the cost-sensitive face recognition, the cost-sensitive parameter settings are presented in Table 1, where $M$ and $I$ are the numbers of gallery subjects and impostor subjects randomly selected from the each database. For each gallery and impostor subject, we randomly selected $N_G$ and $N_I$ samples respectively for granular feature extraction, and the rest samples were used to test the results of image granulation. Since the misclassification costs are only connected with the proportion of the loss functions [10], we set a proportion of the loss functions as presented in Table 1.

Figs. 2 and 3 show the granulated images from the two databases using 2DPCA, 2DLDA and 2DLPP, respectively. As can be seen, the images get clearer with the increasing of the decision steps or granular index $l$, which reflect the image granulation process from rough granule to precise granular. It can be used to represent the incremental available information of an image for sequential three-way decision. At the beginning of the sequential decisions, the available information is insufficient and the image is presented in a rough granular characteristic. With the increasing of available information, the image gets clearer, thus it is presented in a precise granular characteristic. In general, the decision cost will decrease with the increasing of granular index, since the available

<table>
<thead>
<tr>
<th>Database</th>
<th>$M$</th>
<th>$N_G$</th>
<th>$I$</th>
<th>$N_I$</th>
<th>$\lambda_{PN} : \lambda_{BP} : \lambda_{BN} : \lambda_{BP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>15</td>
<td>8</td>
<td>25</td>
<td>8</td>
<td>20:6:4:2</td>
</tr>
<tr>
<td>PIE</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>15</td>
<td>20:6:4:2</td>
</tr>
</tbody>
</table>

Fig. 1. Some samples of AR and PIE face databases used in experiments: (a) 14 samples of one subject from AR face database and (b) 14 samples from one subject (each including 49 samples) in PIE face database.

Fig. 2. Sequential granulation on AR images based on (a) 2DPCA granulation, (b) 2DLDA granulation, and (c) 2DLPP granulation.

Fig. 3. Sequential granulation on PIE images based on (a) 2DPCA granulation, (b) 2DLDA granulation, and (c) 2DLPP granulation.
information is increasing, which can be validated by the following experiments on decision costs and errors.

6.3. Sequential decision costs and errors

In this subsection, we compare the proposed sequential three-way decision with sequential two-way decision in decision costs and errors. In each round of the experiments, we used the same experimental setting as presented in Table 1. We compared sequential three-way and two-way decisions in decision cost (caused by all six recognition decisions defined in Section 3), total cost (decision cost plus test cost), error rate (wrong decision error), high-cost error rate (caused by “false acceptance”, denoted as $\text{err}_{PN}$), low-cost error rate (caused by “false rejection”, denoted as $\text{err}_{NP}$). All error rates were computed as the ratio of misclassification samples count to total samples count. It is worthy noting that the boundary region of three-way decision is viewed as misclassification in our experiments to present a fair comparison with sequential two-way decision. As mentioned in Section 1, if current rough granular images are not precise enough, we may make a boundary decision to collect more information for further certain decision. It may consume some test costs to acquire new high-quality information in such three-way decision process. Normally, the more precise of the granular feature, the higher test cost it consumes, i.e., the test cost is an increasing function of the decision steps and granular index. Therefore, we take an increasing quadratic function $0.001 \cdot l^2$ to represent the test costs. This is only for simplicity in the exposition, and the test cost function can be extended easily to the more general cases. As derived in (3), the optimal sequential three way decision is determined by the thresholds $\alpha, \beta$ and conditional probability $\Pr(P|B)$. The thresholds $\alpha, \beta$ can be computed from (5), and the conditional probability can be

---

Fig. 4. Decision costs and total costs comparison between sequential two-way decision and sequential three-way decision: AR database.

Fig. 5. Decision costs and total costs comparison between sequential two-way decision and sequential three-way decision: PIE database.
Three-way decision. In the case that available information for precise decision is insufficient, the boundary decision presents an optimal choice which leads to a lower decision cost than sequential two-way decision did. This presents a well explanation on how three-way decision presented lower decision costs in comparison with two-way decision. However, we can observe that sequential three-way decision achieved the same lower decision costs and errors were recorded. In the experiment, we used two parameters \( k \) and \( \sigma \) are set as \( k = 5 \) and \( \sigma = 10 \), which are used in constructing a neighborhood graph and similarity matrix for 2DLPP [50,6]. The experimental results about decision costs and total costs on two databases are presented in Figs. 4 and 5, respectively.

As shown in Figs. 4 and 5, both sequential two-way decision and sequential three-way decision achieved decreasing decision costs with the increasing of decision steps. The reason is that the available information for decision is increasing in the sequential decision process, and the decisions get more precise, which leads to the lower decision costs. However, we can see that the sequential three-way decision achieved the lower decision costs than sequential two-way decision did. This is because the boundary decisions are incorporated into three-way decision. In the case that available information for precise decision in insufficient, the boundary decision presents an optimal choice which leads to a lower decision cost. Figs. 4 and 5 also show the comparison results on total costs, which include both decision costs and test costs.

As can be seen, the total costs of the two methods decreased at first and then increased. This happened because the decision costs decreased with the increasing of available information, however, the test costs will increase since acquiring new more information may consume more test cost. At the beginning, the decreasing velocity of decision cost is greater than the increasing velocity of test cost. Thus the total costs decreased. While the decision steps increasing, the decreasing velocity of decision cost is less than the increasing velocity of test cost. Hence, the total costs increased after some certain decision steps. We can see that the minimum total cost was achieved in some middle decision step, which are highlighted in Figs. 4 and 5. It is shown that sequential three-way decision achieved the lower minimum total cost than sequential two-way decision did, and sequential three-way decision achieved the lower total cost than sequential two-way decision did in the whole process of sequential decisions, which validated the effectiveness of the proposed three-way decision methods.

We also compared the error rates of sequential two-way decision and three-way decision. The comparison results are shown in Fig. 6 and Table 2 (average error rates with variance presented). It can be observed that sequential three-way decision achieved the higher error rates than two-way decision, which validate the mechanism of three-way decision: some uncertain facial images cannot be immediately decided, and they are decided in boundary region, which heighten the error rates. However, the boundary decision may lead to a lower cost, as shown in Figs. 4 and 5. Therefore, we prefer the sequential three-way decision results since it is suitable for the cost-sensitive face recognition problem rather than precision-sensitive face recognition problem. From Table 2, we can observe that three-way decision achieved the same low-cost error rates \( \text{err}_{NP} \) as two-way decision did, while it achieved the lower high-cost error rates \( \text{err}_{PN} \) than two-way decision did. This presents a clear explanation on how three-way decision achieved the lower decision costs, and it also validated the effectiveness of sequential decision in the context of cost-sensitive face recognition.

### Table 2

Comparisons on high-cost error (\( \text{err}_{PN} \)) and low-cost error (\( \text{err}_{NP} \)).

<table>
<thead>
<tr>
<th>Decision methods</th>
<th>Granulations</th>
<th>Sequential two-way decision</th>
<th>Sequential three-way decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2DPCA</td>
<td>2DLDA</td>
<td>2DLPP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2DPCA</td>
<td>2DLDA</td>
</tr>
<tr>
<td>AR</td>
<td>( \text{err}_{PN} ) (%)</td>
<td>6.30 ± 0.58</td>
<td>4.82 ± 0.98</td>
</tr>
<tr>
<td></td>
<td>( \text{err}_{NP} ) (%)</td>
<td>8.74 ± 0.46</td>
<td>5.64 ± 1.11</td>
</tr>
<tr>
<td>PIE</td>
<td>( \text{err}_{PN} ) (%)</td>
<td>3.45 ± 0.98</td>
<td>1.75 ± 0.27</td>
</tr>
<tr>
<td></td>
<td>( \text{err}_{NP} ) (%)</td>
<td>6.72 ± 1.44</td>
<td>3.64 ± 0.23</td>
</tr>
</tbody>
</table>

The lower \( \text{err}_{PN} \) in sequential two-way and three-way decision are bolded.
7. Conclusion

In real-world face recognition applications, the costs of different kinds of misrecognition are usually imbalanced, which should be considered in cost-sensitive classification context. On the other hand, the available image information for decision is usually insufficient, which leads to uncertain decision. It can be more certainly decided with the increasing of available information. Based on these two issues, we propose in this paper a sequential three-way decision for cost-sensitive face recognition. The objective of the proposed methods consider a sequential decision process, and it seeks a low decision cost rather than classification error. To represent the image information in the sequential decision process, we develop three two-dimensional subspace projection methods for images granulation. Experiments validate the effectiveness of the granulation methods and show that sequential three-way decision achieves lower decision costs and total costs than traditional two-way decision do. The proposed sequential three-way decision not only presents a new view on cost-sensitive face recognition, but also an applicable simulation on human vision decision process for face recognition.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive advice. This research was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 71201076, 71171107, 61170105, and 61473157, the Natural Science Foundation of Jiangsu, China (BK2011564), and the Ph.D. Programs Foundation of Ministry of Education of China (20120091120004).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.knosys.2015.07.040.

References


